Chapter 2: Fundamentals of convex analysis

Edoardo Amaldi

DEIB – Politecnico di Milano
edoardo.amaldi@polimi.it

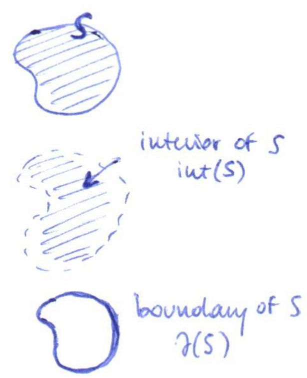
Course material on Beep "[2019-20] - Optimization"

Academic year 2019-20

2.1 Basic concepts

In $\mathbb{R}^n$ with Euclidean norm

- $\underline{x} \in S \subseteq \mathbb{R}^n$ is an **interior point** of S if $\exists\, \varepsilon > 0$ such that $B_\varepsilon(\underline{x}) = \{\underline{y} \in \mathbb{R}^n \ : \ \|\underline{y} - \underline{x}\| < \varepsilon\} \subseteq S$.
- $\underline{x} \in \mathbb{R}^n$ is a **boundary point** of S if, for every $\varepsilon > 0$, $B_\varepsilon(\underline{x})$ contains at least one point of S and one point of its complement $\mathbb{R}^n \setminus S$.
- The set of all the interior points of $S \subseteq \mathbb{R}^n$ is the **interior** of S, denoted by $int(S)$.
- The set of all boundary points of S is the **boundary** of S, denoted by $\partial(S)$.
- $S \subseteq \mathbb{R}^n$ is **open** if $S = int(S)$; S is **closed** if its complement is open. Intuitively, a closed set contains all the points in $\partial(S)$.
- $S \subseteq \mathbb{R}^n$ is **bounded** if $\exists\, M > 0$ such that $\|\underline{x}\| \leq M$ for every $\underline{x} \in S$.
- $S \subseteq \mathbb{R}^n$ closed and bounded is **compact**.



Property:

A set $S \subseteq \mathbb{R}^n$ is closed if and only if every sequence $\{\underline{x}_i\}_{i\in\mathbb{N}} \subseteq S$ that converges, converges to $\underline{x} \in S$.

A set $S \subseteq \mathbb{R}^n$ is compact if and only if every sequence $\{\underline{x}_i\}_{i\in\mathbb{N}} \subseteq S$ admits a subsequence that converges to a point $\underline{x} \in S$.

For preliminaries and convex analysis see:
Bazaraa, Sherali, Shetty, Nonlinear Programming – Theory and Algorithms, third edition, Wiley Interscience, 2006 (Chapters 2 and 3)

Existence of an optimal solution

In general, when minimizing a function $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, we only know that a largest lower bound (infimum) exists, that is

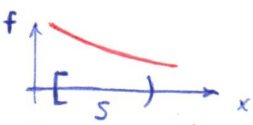
$$\inf_{\underline{x}\in S} f(\underline{x})$$

Theorem (Weierstrass):
Let $S \subseteq \mathbb{R}^n$ be nonempty and compact set, and $f : S \to \mathbb{R}$ be continuous on S. Then there exists a $\underline{x}^* \in S$ such that $f(\underline{x}^*) \leq f(\underline{x})$ for every $\underline{x} \in S$.

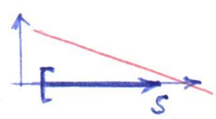
Examples where the result does not hold: S is not closed, S is not bounded or $f(\underline{x})$ is not continuous on S.
(1) (2) (3)

When the problem admits an optimal solution $\underline{x}^* \in S$, we can write $\min_{\underline{x}\in S} f(\underline{x})$.

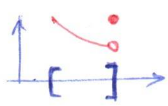
Observation: this result holds in any vector space of finite dimension.

(1)


the optimum ∄ because the endpoint ∉ S

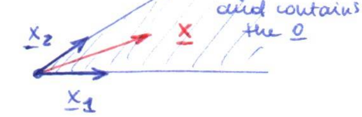
(2)


problem unbounded → ∄ solution

(3)


the minimum is not reached

for example:



$\underline{x} = \alpha_1 \underline{x}_1 + \alpha_2 \underline{x}_2 \quad \alpha_1, \alpha_2 \geq 0$

Cones and affine subspaces

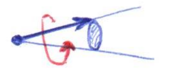
Consider any subset $S \subset \mathbb{R}^n$

Definition: $cone(S)$ denotes the set of all the **conic combinations** of points of S, i.e., all the points $\underline{x} \in \mathbb{R}^n$ such that $\underline{x} = \sum_{i=1}^{m} \alpha_i \underline{x}_i$ with $\underline{x}_1, \ldots, \underline{x}_m \in S$ and $\alpha_i \geq 0$ per every i, $1 \leq i \leq m$.

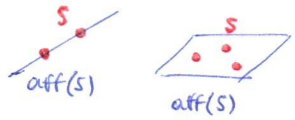
Examples: *polyedral cone* generated by a finite number of vectors, "ice cream" cone in $\mathbb{R}^3$ generated by an infinite number of vectors

Polyedral:

"Ice cream" cone:



for example (2D, 3D):



Definition: $aff(S)$ denotes the smallest **affine subspace** that contains S.

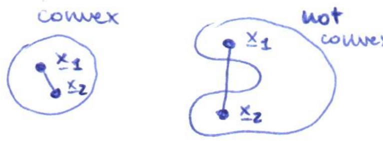
$aff(S)$ coincides with the set of all the **affine combinations** of points in S, i.e., all the points $\underline{x} \in \mathbb{R}^n$ such that $\underline{x} = \sum_{i=1}^{m} \alpha_i \underline{x}_i$ with $\underline{x}_1, \ldots, \underline{x}_m \in S$ and with $\sum_{i=1}^{m} \alpha_i = 1$, where $\alpha_i \in \mathbb{R}$ for every i, $1 \leq i \leq m$.

Examples: straight line containing two points in $\mathbb{R}^2$, plane containing three points in general position in $\mathbb{R}^3$

2.2 Elements of convex analysis

Definition:
A set $C \subset \mathbb{R}^n$ is **convex** if

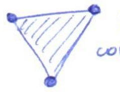
$$\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2 \in C \quad \forall \underline{x}_1, \underline{x}_2 \in C \quad \text{and} \quad \forall \alpha \in [0,1].$$



A point $\underline{x} \in \mathbb{R}^n$ is a **convex combination** of $\underline{x}_1, \ldots, \underline{x}_m \in \mathbb{R}^n$ se

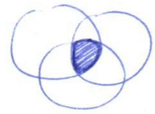
$$\underline{x} = \sum_{i=1}^{m} \alpha_i \underline{x}_i$$

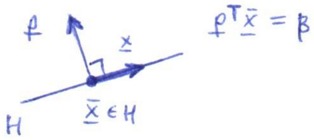
with $\alpha_i \geq 0$ for every i, $1 \leq i \leq m$, and $\sum_{i=1}^{m} \alpha_i = 1$.



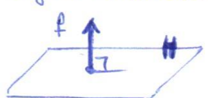
set of all convex combinations

Property: If C_i with $i = 1, \ldots, k$ are convex, then $\cap_{i=1}^{k} C_i$ is convex.





In higher dimension:



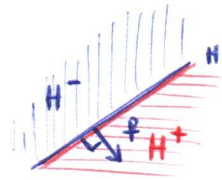
Examples of convex sets

1) Hyperplane $H = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t \underline{x} = \beta\}$ with $\underline{p} \neq \underline{0}$.

For $\underline{\bar{x}} \in H$, $\underline{p}^t \underline{\bar{x}} = \beta$ implies $H = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t(\underline{x} - \underline{\bar{x}}) = 0\}$ and hence $\underline{p}$ is orthogonal to all the vectors $(\underline{x} - \underline{\bar{x}})$ for $\underline{x} \in H$

N.B.: H is closed since $H = \partial(H)$

2) Closed half-spaces $H^+ = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t \underline{x} \geq \beta\}$ and $H^- = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t \underline{x} \leq \beta\}$ with $\underline{p} \neq \underline{0}$.



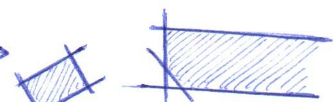
both H^+ and H^- includes H (otherwise they wouldn't be closed)

3) Feasible region $X = \{\underline{x} \in \mathbb{R}^n \ : \ A\underline{x} \geq \underline{b}, \ \underline{x} \geq \underline{0}\}$ of a Linear Program (LP)

$$\begin{array}{ll} \min & \underline{c}^t \underline{x} \\ s.t. & A\underline{x} \geq \underline{b} \\ & \underline{x} \geq \underline{0} \end{array}$$

X is a convex and closed subset (intersection of $m+n$ closed half-spaces if $A \in \mathbb{R}^{m \times n}$).

Definition: The intersection of a finite number of closed half-spaces is a **polyedron.**

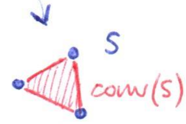


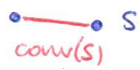
finite / infinite

N.B.: The set of optimal solutions of a LP is a polyhedron (just add $\underline{c}^t \underline{x} = z^*$ with z^* optimal value) and hence convex.

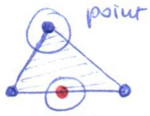
Definition: The **convex hull** of $S \subseteq \mathbb{R}^n$, denoted by $conv(S)$, is the intersection of all convex sets containing S.

$conv(S)$ coincides with the set of all convex combinations of points in S.
Two equivalent characterizations (external/internal descriptions).





extreme point



not extreme point

Definition: Given a convex set C of $\mathbb{R}^n$, $\underline{x} \in C$ is an **extreme point** of C if it cannot be expressed as convex combination of two different points of C, that is

$$\underline{x} = \alpha \underline{x}_1 + (1-\alpha)\underline{x}_2 \quad \text{with } \underline{x}_1, \underline{x}_2 \in C \text{ and } \alpha \in (0,1)$$

implies that $\underline{x}_1 = \underline{x}_2$.

Examples: convex sets with a finite and infinite number of extreme points.

all the points on the boundary are extreme points (can't be expressed through other points)

Projection on a convex set

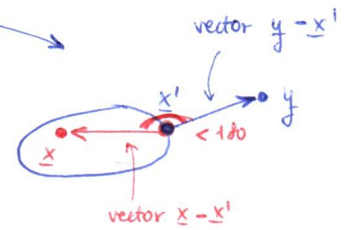
Projection of a point on a convex set to which it does not belong to.

Lemma (Projection):
Let $C \subseteq \mathbb{R}^n$ be a nonempty, closed and convex set, then for every $\underline{y} \notin C$ there exists a unique $\underline{x}' \in C$ at minimum distance from $\underline{y}$.
Moreover, $\underline{x}' \in C$ is the closest point to $\underline{y}$ if and only if

$$\boxed{(\underline{y} - \underline{x}')^t(\underline{x} - \underline{x}') \leq \underline{0} \quad \forall \underline{x} \in C.}$$

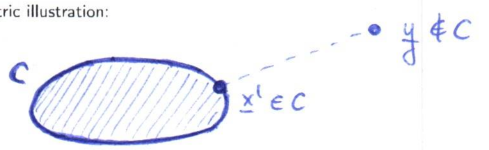
(characterization of $\underline{x}'$)

we're saying that the projections on the line connecting y and $\underline{x}$ have opposite directions:



Definition: $\underline{x}'$ is the **projection** of $\underline{y}$ on C. (orthogonal projection)

Geometric illustration:



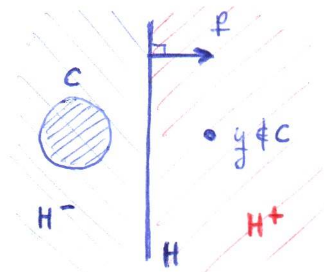
Separation theorem and consequences

Geometrically intuitive but fundamental result.

Theorem (Separating hyperplane)
Let $C \subset \mathbb{R}^n$ be a nonempty, closed and convex set and $\underline{y} \notin C$, then there exists $\underline{p} \in \mathbb{R}^n$ such that $\underline{p}^t\underline{x} < \underline{p}^t\underline{y}$ for every $\underline{x} \in C$.

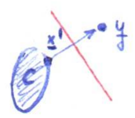
Thus there exists a hyperplane $H = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t\underline{x} = \beta\}$ with $\underline{p} \neq \underline{0}$ that separates $\underline{y}$ from C, i.e., such that

$$C \subseteq H^- = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t\underline{x} \leq \beta\} \quad \text{and} \quad \underline{y} \notin H^- \quad (\underline{p}^t\underline{y} > \beta)$$

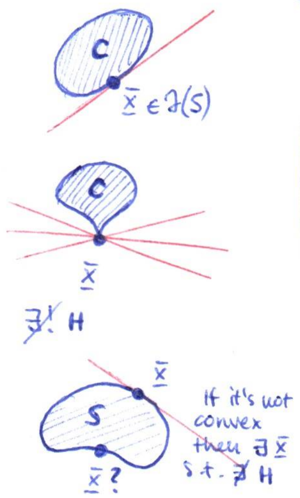


$\exists \underline{p} \in \mathbb{R}^n, \underline{p} \neq \underline{0}$:
$\exists H$: $C \subseteq H^-$ and $y \in H^+$
$(\underline{p}^T\underline{x} < \underline{p}^T y \quad \forall \underline{x} \in C)$

Geometric illustration:

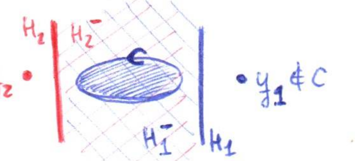


Proof: According to the Lemma, $(\underline{y} - \underline{x}')^t(\underline{x} - \underline{x}') \leq \underline{0} \quad \forall \underline{x} \in C$.
Taking $\underline{p} = (\underline{y} - \underline{x}') \neq \underline{0}$ and $\beta = (\underline{y} - \underline{x}')^t\underline{x}'$ we have $\underline{p}^t\underline{x} \leq \beta \ \forall \underline{x} \in C$,
while $\underline{p}^t\underline{y} - \beta = (\underline{y} - \underline{x}')^t(\underline{y} - \underline{x}') = \|\underline{y} - \underline{x}'\|^2 > 0$ since $\underline{y} \notin C$. □



Three important **consequences**:

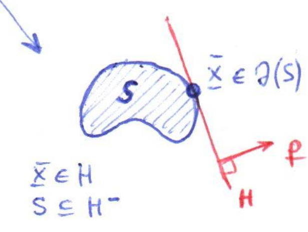
1) Any nonempty, closed and convex set $C \subseteq \mathbb{R}^n$ is the intersection of all closed half-spaces containing it.



not necessarily convex (otherwise := C)

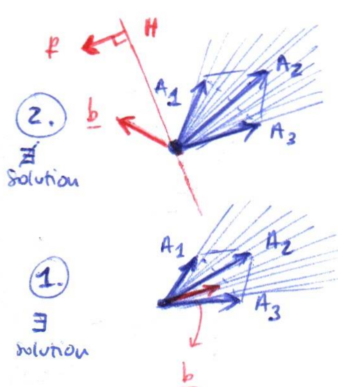
Definition: Let $S \subseteq \mathbb{R}^n$ be nonempty and $\underline{\bar{x}} \in \partial(S)$ (boundary w.r.t. $aff(S)$), $H = \{\underline{x} \in \mathbb{R}^n \ : \ \underline{p}^t(\underline{x} - \underline{\bar{x}}) = 0\}$ is a **supporting hyperplane** of S at $\underline{\bar{x}}$ if $S \subseteq H^-$ or $S \subseteq H^+$.

2) Supporting hyperplane:
If $C \neq \emptyset$ is a convex set in $\mathbb{R}^n$, then for every $\underline{\bar{x}} \in \partial(C)$ there exists a supporting hyperplane H at $\underline{\bar{x}}$, i.e., $\exists \ \underline{p} \neq \underline{0}$ such that $\underline{p}^t(\underline{x} - \underline{\bar{x}}) \leq 0$, for each $\underline{x} \in C$.



A convex set admits at least a supporting hyperplane at each boundary point. For nonconvex sets, such a hyperplane may not exist.

Examples: cases with $1/\infty/0$ supporting hyperplanes at a given boundary point $\underline{\bar{x}}$



Central result of Optimization (also of Game theory) from which we will derive the optimality conditions for Nonlinear Optimization.

3) Farkas Lemma:
Let $A \in \mathbb{R}^{m \times n}$ and $\underline{b} \in \mathbb{R}^m$. Then

$$\overset{1.}{\boxed{\exists\, \underline{x} \in \mathbb{R}^n \text{ such that } A\underline{x} = \underline{b} \text{ and } \underline{x} \geq \underline{0}}} \Leftrightarrow \overset{2.}{\boxed{\nexists\, \underline{y} \in \mathbb{R}^m \text{ such that } \underline{y}^tA \leq \underline{0}^t \text{ and } \underline{y}^t\underline{b} > 0.}}$$

one system admits a solution ⟺ another system doesn't admit a solution

$$\left.\begin{array}{l}\exists \underline{x} \in \mathbb{R}^n: \\ A\underline{x} = \underline{b} \\ \underline{x} \geq \underline{0}\end{array}\right\} \Leftrightarrow \left\{\begin{array}{l}\nexists y \in \mathbb{R}^m: \\ y^TA \leq \underline{0}^T \\ y^T\underline{b} > 0\end{array}\right.$$

In this form, it provides an infeasibility certificate for a given linear system, but it is also known as the theorem of the alternative.

Alternative: exactly one of the two systems $A\underline{x} = \underline{b}, \underline{x} \geq \underline{0}$ and $\underline{y}^tA \leq \underline{0}^t, \underline{y}^t\underline{b} > 0$ is feasible.

Geometric interpretation:

$\underline{b}$ belongs to the convex cone generated by the columns $A_1, \ldots, A_n$ of A, i.e., to $cone(A) = \{\underline{z} \in \mathbb{R}^m \ : \ \underline{z} = \sum_{j=1}^n \alpha_j A_j, \alpha_1 \geq 0, \ldots, \alpha_n \geq 0\}$, if and only if no hyperplane separating $\underline{b}$ from $cone(A)$ exists.

Alternative: $\overset{1.}{\boxed{\underline{b} \in cone(A)}}$ or $\overset{2.}{\boxed{\underline{b} \notin cone(A)}}$ (hence $\exists$ hyperplane separating $\underline{b}$ from $cone(A)$)

either $\underline{b} \in / \notin cone(A)$:
- $\in \Rightarrow A\underline{x} = \underline{b}$ has a solution
- $\notin \Rightarrow \exists H$ hyperplane that separate $\underline{b}$ from cone(A)

Proof:

($\Rightarrow$) Consider $\underline{\tilde{x}} \geq \underline{0}$ s.t. $A\underline{\tilde{x}} = \underline{b}$ ($\equiv$ 1. is feasible). For all y s.t. $y^T A \leq \underline{0}$ we have: $y^T \underline{b} = y^T A \underline{\tilde{x}} \leq 0$
$\Rightarrow \nexists\, y$ s.t. $y^T A \underline{x} \leq 0$ and $y^T \underline{b} > 0$

($\Leftarrow$) Assume that $A\underline{x} = \underline{b}$, $\underline{x} \geq \underline{0}$ is unfeasible, i.e. $\underline{b} \notin \text{cone}(A) = \{\underline{z} \in \mathbb{R}^m : \underline{z} = \sum_{j=1}^n \alpha_j A^j,\ \alpha_j \geq 0\ \forall j\}$
(cone(A) : $\neq \emptyset$, closed, convex). $b \notin \text{cone}(A) \Rightarrow$ we apply hyperplane separating theorem:
$\Rightarrow \exists\, \underline{p} \in \mathbb{R}^m$ and $\beta \in \mathbb{R}$ s.t. $(\underline{p}^T \underline{b} > \beta) \wedge (\underline{p}^T \underline{z} \leq \beta)$ $\forall \underline{z} \in \text{cone}(A)$
$\Rightarrow \text{cone}(A) \in H^-$ $(H^- = \{\ldots\ \underline{p}^T \underline{x} \leq \beta\})$
since $\underline{0} \in \text{cone}(A)$, $\beta \geq 0$ and hence $\underline{p}^T \underline{b} > 0$. Moreover: $\underline{p}^T \underline{z} \leq \beta$ $\forall \underline{z} \in \text{cone}(A)$
$\Rightarrow \underline{p}^T A \underline{x} \leq \beta$ $\forall \underline{x} \geq 0$ and since $\underline{x} \geq \underline{0} \Rightarrow \underline{p}^T A \leq \underline{0}$
Thus $\underline{p} \neq \underline{0}$, $\underline{p}^T A \leq \underline{0}$ and $\underline{p}^T \underline{b} > 0 \Rightarrow$ system 2. is feasible, $y = \underline{p}$
(we proved $\bar{1}. \to \bar{2}.$)

Application of Farkas Lemma: asset pricing in absence of arbitrage

Single period, m assets are traded, n possible states (scenarios).

Assets can be bought or sold "short", i.e., with the promise to buy them back at the end.

Portfolio $\underline{y} \in \mathbb{R}^m$, with y_i = amount invested in asset i.
A negative value of y_i indicates a "short" position.

If p_i is the price of a unit of asset i at the beginning, the portfolio cost is: $\underline{p}^t \underline{y}$

Consider $A \in \mathbb{R}^{m \times n}$ with a_{ij} = payoff of one euro invested in asset i if state j occurs.

At the end all assets are sold (we receive a payoff of $a_{ij} y_i$) and the "short" positions are covered (we pay $a_{ij}|y_i|$ and hence have a payoff of $a_{ij} y_i$).

Portfolio value: $\underline{v}^t = \underline{y}^t A$

Absence of arbitrage condition: any portfolio with nonnegative values for all states must have a nonnegative cost.

Algebraically: $\nexists\, \underline{y}$ such that $\underline{y}^t A \geq \underline{0}^t$ and $\underline{p}^t \underline{y} < 0$.

According to Farkas Lemma:

No possibility of arbitrage exist ($\nexists\, \underline{y}$ such that $\underline{y}^t A \geq \underline{0}^t$ and $\underline{p}^t \underline{y} < 0$)

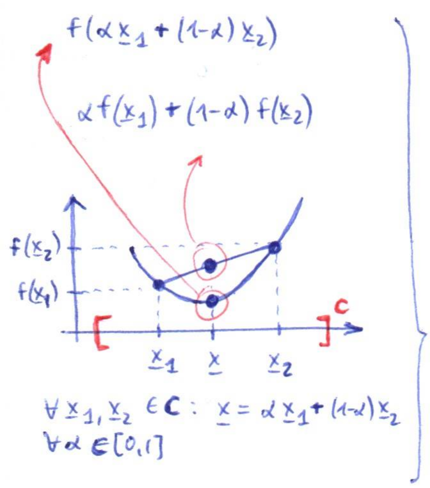
if and only if

$\exists\, \underline{q} \in \mathbb{R}^n$ with $\underline{q} \geq \underline{0}$ such that the asset price vector $\underline{p}$ satisfies

$$A\underline{q} = \underline{p}.$$

If the market is efficient, some "state prices" q_i exist.

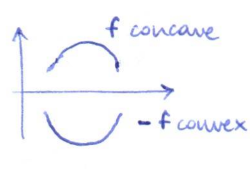
N.B.: in general $\underline{q}$ is not unique.



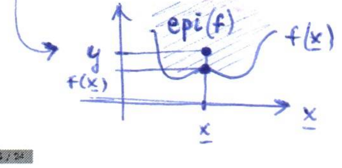
2.2.2 Convex functions

Definitions:

- A function $f : C \to \mathbb{R}$ defined on a convex set $C \subseteq \mathbb{R}^n$ is **convex** if
  $f(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2) \leq \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2)$ $\forall \underline{x}_1, \underline{x}_2 \in C$ and $\forall \alpha \in [0,1]$,
- f is **strictly convex** if the inequality holds with $<$ for $\forall \underline{x}_1, \underline{x}_2 \in C$ with $\underline{x}_1 \neq \underline{x}_2$ and $\forall \alpha \in (0,1)$.
- is **concave** if $-f$ is convex; f is **linear** if it is both convex and concave.



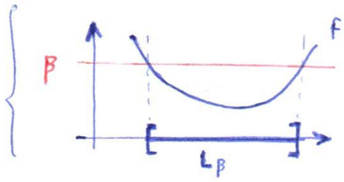
- The **epigraph** of $f : S \subseteq \mathbb{R}^n \to \mathbb{R}$, denoted by $epi(f)$, is the subset of $\mathbb{R}^{n+1}$
  $$epi(f) = \{(\underline{x}, y) \in S \times \mathbb{R} \ :\ f(\underline{x}) \leq y\}.$$



- Let $f : C \to \mathbb{R}$ be convex, the **domain** of f is the subset of $\mathbb{R}^n$
  $$dom(f) = \{\underline{x} \in C \ :\ f(\underline{x}) < +\infty\}.$$

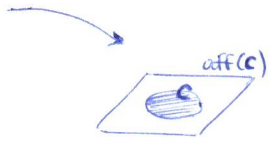
Property:

Let $C \subseteq \mathbb{R}^n$ be a (nonempty) convex set and $f : C \to \mathbb{R}$ be a convex function.



- For each real β (also for $+\infty$), the *level set*
$$L_\beta = \{\underline{x} \in C : f(\underline{x}) \leq \beta\} \quad \text{and} \quad \{\underline{x} \in C : f(\underline{x}) < \beta\}$$
is a convex subset of $\mathbb{R}^n$.

- The function f is continuous in the relative interior (with respect to *aff*(C)) of its domain.



- The function f is convex if and only if *epi*(f) is a convex subset of $\mathbb{R}^{n+1}$ (exercise 1.3).

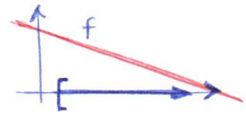
Optimal solution of convex problems

Consider $\min_{\underline{x} \in C \subseteq \mathbb{R}^n} f(\underline{x})$ where $C \subseteq \mathbb{R}^n$ is a convex set and f is a convex function.

Proposition:

i) If $C \subseteq \mathbb{R}^n$ is convex and $f : C \to \mathbb{R}$ is convex, each local minimum of f on C is a global minimum.

ii) If f is strictly convex on C, then there exists at most one global minimum (the problem may be unbounded).



Proof: Suppose $\underline{x}'$ is a local minimum and $\exists\ \underline{x}^* \in C$ such that $f(\underline{x}^*) < f(\underline{x}')$.

i) Since f is convex
$$f(\alpha\underline{x}' + (1-\alpha)\underline{x}^*) \leq \alpha f(\underline{x}') + (1-\alpha) f(\underline{x}^*) < f(\underline{x}') \quad \forall \alpha \in (0,1)$$
contradicts the fact that $\underline{x}'$ is a local minimum.

ii) If f is strictly convex and $\underline{x}_1^*$ and $\underline{x}_2^*$ are two global minima, the convexity of C implies
$$\frac{1}{2}(\underline{x}_1^* + \underline{x}_2^*) \in C$$
and strict convexity of f implies
$$f(\frac{1}{2}(\underline{x}_1^* + \underline{x}_2^*)) < \frac{1}{2} f(\underline{x}_1^*) + \frac{1}{2} f(\underline{x}_2^*).$$
Thus $\underline{x}_1^*$ and $\underline{x}_2^*$ cannot be two global minima. □

Special case: linear programming problems

Consider any linear program (LP)

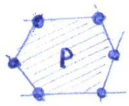
$$\begin{array}{ll} \min & \underline{c}^t \underline{x} \\ s.t. & A\underline{x} \geq \underline{b} \\ & \underline{x} \geq \underline{0} \end{array} \Bigg\} \ P = \text{non empty feasible region}$$

where P denotes the nonempty feasible region, polyhedron.

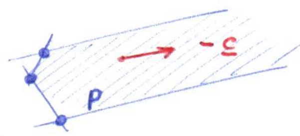
Proposition:
Given any LP with a nonempty feasible region, then either there exists (at least) one optimal extreme point or the value of the objective function is unbounded below on the feasible region.

≡ in the special case of linear programming we can focus on the extreme points

Geometric illustration:



either the solution is one of the extreme points, or:

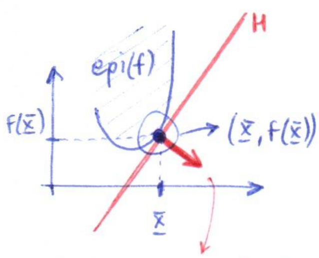


Characterizations of convex functions

1) Proposition: A continuously differentiable function (of class C^1) $f : C \to \mathbb{R}$ defined on an open and nonempty convex set $C \subseteq \mathbb{R}^n$ is convex if and only if
$$f(\underline{x}) \geq f(\overline{\underline{x}}) + \nabla^t f(\overline{\underline{x}})(\underline{x} - \overline{\underline{x}}) \qquad \forall \underline{x}, \overline{\underline{x}} \in C.$$

first order taylor expansion of the function at $\overline{\underline{x}}$

f is strictly convex if and only if the inequality holds with $>$ for every pair $\underline{x}, \overline{\underline{x}} \in C$ with $\underline{x} \neq \overline{\underline{x}}$.



this normal vector is: $\begin{bmatrix} \nabla^T f(\overline{x}) \\ -1 \end{bmatrix}$

$y = f(\overline{x}) + \nabla^T f(\overline{x})(x - \overline{x})$
$= f(\overline{x}) + \nabla^T f(\overline{x})x - \nabla^T f(\overline{x})\overline{x}$

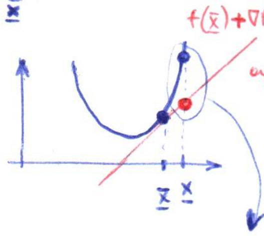
$\boxed{\nabla^T f(\overline{x})x - y = -f(\overline{x}) + \nabla^T f(\overline{x})\overline{x}}$

Definition: *Directional derivative* of f: $\lim_{\alpha \to 0^+} \frac{f(\overline{\underline{x}} + \alpha(\underline{x} - \overline{\underline{x}})) - f(\overline{\underline{x}})}{\alpha} = \nabla^t f(\overline{\underline{x}})(\underline{x} - \overline{\underline{x}})$

Geometric interpretation:

The linear approximation of f at $\overline{\underline{x}}$ (1st order Taylor's expansion) bounds below $f(\underline{x})$ and
$$H = \{\begin{pmatrix} \underline{x} \\ y \end{pmatrix} \in \mathbb{R}^{n+1} : (\nabla^t f(\overline{\underline{x}}) \quad -1) \begin{pmatrix} \underline{x} \\ y \end{pmatrix} = -f(\overline{\underline{x}}) + \nabla^t f(\overline{\underline{x}})\ \overline{\underline{x}}\ \}$$
is a supporting hyperplane of *epi*(f) at $(\overline{\underline{x}}, f(\overline{\underline{x}}))$, with *epi*$(f) \subseteq H^-$.

$f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x})$ linear approxim. at $\overline{x}$



$f(x) \geq f(\overline{x}) + \nabla^T f(\overline{x})(x - \overline{x})$

We can use $f(\overline{x}) + \nabla f(\overline{x})(x - \overline{x})$ as an hyperplane

2) Proposition: A twice continuously differentiable function (of class C^2) $f : C \to \mathbb{R}$ defined on an open and nonempty convex set $C \subseteq \mathbb{R}^n$ is convex if and only if the Hessian matrix $\nabla^2 f(\underline{x}) = (\frac{\partial^2 f}{\partial x_i \partial x_j})$ is positive semidefinite at every $\underline{x} \in C$.

For C^2 functions, if $\nabla^2 f(\underline{x})$ is positive definite $\forall \underline{x} \in C$ then $f(\underline{x})$ is strictly convex.

N.B.: This condition is sufficient but not necessary: $f(x) = x^4$ is strictly convex but $f''(0) = 0$.

Definition:
A symmetric matrix A $n \times n$ is *positive definite* if $\underline{y}^t A \underline{y} > 0 \quad \forall \underline{y} \in \mathbb{R}^n$ with $\underline{y} \neq \underline{0}$.
A symmetric matrix A $n \times n$ is *positive semidefinite* if $\underline{y}^t A \underline{y} \geq 0 \quad \forall \underline{y} \in \mathbb{R}^n$.

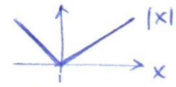
We can consider the eigenvalues: if they're strictly positive then the matrix is positive def., if they're ≥ 0 then the matrix is positive semidef.

Equivalent definitions: based on the sign of the eigenvalues/principal minors of A or of the diagonal coefficients of specific factorizations of A (e.g., Cholesky factorization).

$A = LL^T$ $\quad L =$ lower triangular matrix

Subgradient of convex and concave functions

Convex/concave (continuous) functions that are not everywhere differentiable, e.g. $f(x) = |x|$.



C^1 except for $x = 0$

Generalization of the concept of gradient for C^1 functions to piecewise C^1 functions.

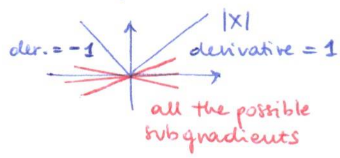
Definitions: Let $C \subseteq \mathbb{R}^n$ be a convex set and $f : C \to \mathbb{R}$ a convex function on C

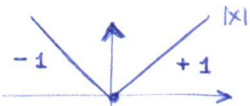
- a vector $\underline{\gamma} \in \mathbb{R}^n$ is a **subgradient** of f at $\overline{x} \in C$ if

$$\boxed{f(\underline{x}) \geq f(\overline{x}) + \underline{\gamma}^t(\underline{x} - \overline{x}) \qquad \forall \underline{x} \in C,} \quad (*)$$

- the **subdifferential**, denoted by $\partial f(\underline{x})$, is the set of all the subgradients of f at $\underline{x}$.

Example:





- if $x > 0$: $\partial f(x) = \{1\}$
- if $x < 0$: $\partial f(x) = \{-1\}$
- $\partial f(0) = [-1, 1]$

∞ points since $x = 0$ is not differentiable

Example: For $f(x) = x^2$, in $\overline{x} = 3$ the only subgradient is $\gamma = 6$. ($f'(x) = 2x$, $f'(3) = 6$)
Indeed, $0 \leq (x-3)^2 = x^2 - 6x + 9$ implies for every x:

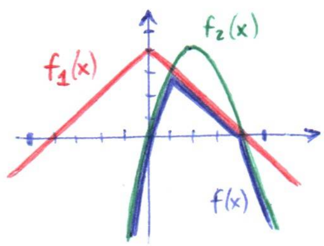
$$f(x) = x^2 \geq 6x - 9 = 9 + 6(x-3) = f(\overline{x}) + 6(x - \overline{x})$$

(we wanted to check $(*)$)

Other examples:

1) For $f(x) = |x|$ it is clear that: $\gamma = 1$ if $x > 0$, $\gamma = -1$ if $x < 0$, $\partial f(x) = [-1, 1]$ if $x = 0$.

2) Consider $f(x) = \min\{f_1(x), f_2(x)\}$ with $f_1(x) = 4 - |x|$ and $f_2(x) = 4 - (x-2)^2$.



Since $f_2(x) \geq f_1(x)$ for $1 \leq x \leq 4$,

$$f(x) = \begin{cases} 4 - x & 1 \leq x \leq 4 \\ 4 - (x-2)^2 & \text{otherwise} \end{cases} \quad \begin{matrix} \longrightarrow \partial : -1 \\ \longrightarrow \partial : -2(x-2) \end{matrix}$$

which is concave.

concave and not differentiable (C^1) since there are 2 points s.t. they're not differentiable

$\gamma = -1$ for $x \in (1, 4)$,
$\gamma = -2(x-2)$ for $x < 1$ or $x > 4$,
$\gamma \in [-1, 2]$ at $x = 1$, $\longrightarrow \begin{cases} \partial : -1 \\ \partial : -2(x-2) = +2 \end{cases} \Rightarrow [-1, +2]$
$\gamma \in [-4, -1]$ at $x = 4$. $\longrightarrow \begin{cases} \partial : -1 \\ \partial : -2(x-2) = -4 \end{cases} \Rightarrow [-4, -1]$

Properties:

1) A convex function $f : C \to \mathbb{R}$ admits at least a subgradient at every interior point $\overline{x}$ of C. In particular, if $\overline{x} \in int(C)$ then there exists $\underline{\gamma} \in \mathbb{R}^n$ such that

$$H = \{(\underline{x}, y) \in \mathbb{R}^{n+1} \ : \ y = f(\overline{x}) + \underline{\gamma}^t(\underline{x} - \overline{x})\}$$

is a supporting hyperplane of $epi(f)$ at $(\overline{x}, f(\overline{x}))$.

N.B.: The existence of (at least) a subgradient at every point of $int(C)$, with C convex, is a necessary and sufficient condition for f to be convex on $int(C)$.

2) If f is a convex function and $\underline{x} \in C$, $\partial f(\underline{x})$ is a nonempty, convex, closed and bounded set.

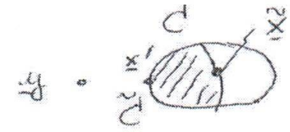
at least the gradient

3) $\underline{x}^*$ is a (global) minimum of a convex function $f : C \to \mathbb{R}$ if and only if $\underline{0} \in \partial f(\underline{x}^*)$.

$f(\underline{x}) \geq f(\underline{x}^*) + \gamma^T(\underline{x} - \underline{x}^*) \quad \forall \underline{x} \in C$
$\gamma^T = 0 \Rightarrow f(\underline{x}) \geq f(\underline{x}^*) \quad \forall \underline{x} \in C$

$\Rightarrow$ global minimum

• Existence



$$C \neq \emptyset \Rightarrow \exists\, \tilde{x} \in C \text{ and } \tilde{C} = C \cap \{ x \in \mathbb{R}^n : \|y - x\| \leq \|y - \tilde{x}\| \}$$

compact

Since $d(y, C) = \inf \{ \|y - x\| : x \in C \} = \inf \{ \|y - x\| : x \in \tilde{C} \}$

and $\|y - \cdot\|$ continuous and $\tilde{C}$ is compact,

Weierstrass th. $\Rightarrow \exists\, x'$ closest to y.

Uniqueness because distance is strictly convex.

• Sufficient ($\Leftarrow$) $\qquad \forall x \in C$

$\|a \mp b\|^2 = \|a\|^2 + \|b\|^2 \mp 2 a^t b$

$$\|y - x\|^2 = \|y - x' + x' - x\|^2 = \|y - x'\|^2 + \underbrace{\|x' - x\|^2}_{\geq 0} - 2 \underbrace{(y - x')^t (x' - x)}_{\substack{\geq 0 \\ \text{assumption}}}$$

$$\Rightarrow \|y - x\|^2 \geq \|y - x'\|^2 \quad \forall x \in C$$

that is, x' is closest

• Necessary ($\Rightarrow$)

x' closest i.e. $\|y - x\|^2 \geq \|y - x'\|^2 \quad \forall x \in C$

By convexity $\qquad x' + \alpha(x - x') \in C \quad \forall x \in C, \forall \alpha \in [0, 1]$

$$\Rightarrow \|y - x' - \alpha(x - x')\|^2 \geq \|y - x'\|^2$$

Moreover

$$\|y - x' - \alpha(x - x')\|^2 = \|y - x'\|^2 + \underbrace{\alpha^2 \|x - x'\|^2 - 2\alpha (y - x')^t (x - x')}_{\Rightarrow \ \geq 0}$$

$$\Rightarrow \quad 2 \cancel{\alpha} (y - x')^t (x - x') \leq \alpha^{\cancel{2}} \|x - x'\|^2$$

≤ 0 letting $\alpha \to 0^+$

□

CONVEX ANALYSIS

WEIERSTRASS THEOREM

$$\left.\begin{array}{l} S \subseteq \mathbb{R}^n \text{ non-empty and compact} \\ f: S \to \mathbb{R} \text{ continuous on } S \end{array}\right\} \Longrightarrow \exists \underline{x}^* \in S: \quad f(\underline{x}^*) \leq f(\underline{x}) \quad \forall \underline{x} \in S$$

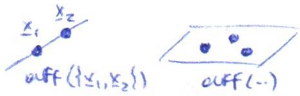
It doesn't hold if: S is not closed, S is not bounded, f is not continuous on S.

- Conic combination

  $\text{cone}(S) = \{ \underline{x} \in \mathbb{R}^n : \quad \underline{x} = \sum_{i=1}^m \alpha_i \underline{x}_i \quad \underline{x}_i \in S, \quad \alpha_i \geq 0 \}$

- Affine combinations

  $\text{aff}(S) = \{ \underline{x} \in \mathbb{R}^n : \quad \underline{x} = \sum_{i=1}^m \alpha_i \underline{x}_i \quad \underline{x}_i \in S, \quad \alpha_i \in \mathbb{R} : \quad \sum_{i=1}^m \alpha_i = 1 \}$



- Convex combinations

  $\underline{x} \in \mathbb{R}^n : \quad \underline{x} = \sum_{i=1}^m \alpha_i \underline{x}_i \quad \underline{x}_i \in \mathbb{R}^m, \quad \alpha_i \geq 0 : \quad \sum_{i=1}^m \alpha_i = 1$

- Hyperplane

  $H = \{ \underline{x} \in \mathbb{R}^n : \quad \underline{p}^T \underline{x} = \beta \} \qquad \underline{p} \neq \underline{0}$

  $H^+ = \{ \underline{x} \in \mathbb{R}^n : \quad \underline{p}^T \underline{x} \geq \beta \}$

  $H^- = \{ \underline{x} \in \mathbb{R}^n : \quad \underline{p}^T \underline{x} \leq \beta \}$

- Convex hull

  $\text{conv}(S)$ = intersection of all the convex sets containing $S \subseteq \mathbb{R}^n$

- Extreme point

  $C \subseteq \mathbb{R}^n$ convex set. $\underline{x} \in C$ is an extreme point if it cannot be expressed as a convex combination of two different points of C (i.e. $\underline{x} = \alpha \underline{x}_1 + (1-\alpha)\underline{x}_2 \quad \underline{x}_{1,2} \in C$)

PROJECTION LEMMA

$C \subseteq \mathbb{R}^n$ non empty, closed, convex $\Longrightarrow$

- $\forall \underline{y} \notin C \quad \exists! \, \underline{x}^* \in C$ at min dist from $\underline{y}$
- $\underline{x}^*$ is the closest to $\underline{y} \iff (\underline{y} - \underline{x}^*)^T (\underline{x} - \underline{x}^*) \leq 0 \quad \forall \underline{x}$

SEPARATING HYPERPLANE

$$\left.\begin{array}{l} C \subseteq \mathbb{R}^n \text{ non empty, closed, convex} \\ \underline{y} \notin C \end{array}\right\} \Longrightarrow \exists \underline{p} \in \mathbb{R}^n : \quad \underline{p}^T \underline{x} \leq \underline{p}^T \underline{y} \quad \forall \underline{x} \in C$$

and so $\exists H = \{ \underline{x} \in \mathbb{R}^n : \underline{p}^T \underline{x} = \beta \}$ that separates $\underline{y}$ from C

(i.e. $C \subseteq H^- = \{ \underline{x} \in \mathbb{R}^n : \underline{p}^T \underline{x} \leq \beta \}, \quad \underline{y} \notin H^-$)

- Supporting hyperplane

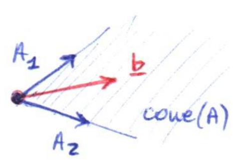
  $S \subseteq \mathbb{R}^n$ non-empty, $\bar{\underline{x}} \in \partial(S)$: $H = \{ \underline{x} \in \mathbb{R}^n : \underline{p}^T (\underline{x} - \bar{\underline{x}}) = 0 \}$ is a supporting hyperplane of S at $\bar{\underline{x}}$ if $S \subseteq H^+$ or $S \subseteq H^-$.

  Remark: a convex set admits at least a supporting hyperplane at each boundary point (i.e. $\forall \bar{\underline{x}} \in \partial(C) \quad \exists \underline{p} \neq \underline{0} : \quad \underline{p}^T (\underline{x} - \bar{\underline{x}}) \leq 0 \quad \forall \underline{x} \in C$)

FARKAS LEMMA

$$\left.\begin{array}{l} A \in \mathbb{R}^{m \times n} \\ \underline{b} \in \mathbb{R}^m \end{array}\right\} \Longrightarrow \left\{ \exists \underline{x} \in \mathbb{R}^n : \begin{array}{l} A\underline{x} = \underline{b} \\ \underline{x} \geq \underline{0} \end{array} \right. \iff \left\{ \nexists \underline{y} \in \mathbb{R}^m : \begin{array}{l} \underline{y}^T A \leq \underline{0}^T \\ \underline{y}^T \underline{b} > 0 \end{array} \right.$$

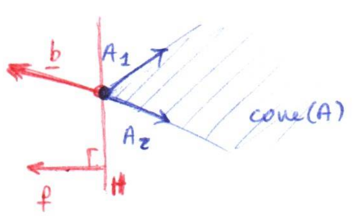
only one system admits solution



A_j = j-th column of A

$\underline{b} \in \text{cone}(A)$

$A\underline{x} = \underline{b}$ has a solution



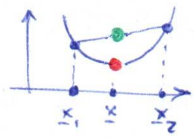
$\underline{b} \notin \text{cone}(A)$

$\exists H$ hyperplane that separates $\underline{b}$ from cone(A)

- Convex function

  $C \subseteq \mathbb{R}^n$ convex, $f: C \to \mathbb{R}$ convex if:

  $$f(\alpha \underline{x}_1 + (1-\alpha)\underline{x}_2) \leq \alpha f(\underline{x}_1) + (1-\alpha) f(\underline{x}_2) \qquad \forall \underline{x}_1, \underline{x}_2 \in C, \quad \alpha \in [0,1]$$



- Epigraph of f

  $f: S \subseteq \mathbb{R}^n \to \mathbb{R}, \qquad \mathrm{epi}(f) = \{ (\underline{x}, y) \in S \times \mathbb{R} : f(\underline{x}) \leq y \}$

CHARACT. OF CONVEX FUNCTIONS

- $f: C \to \mathbb{R}$, $f \in C^1(C)$, $C \subseteq \mathbb{R}^n$ open, non-empty convex $\Longrightarrow$ f convex $\Longleftrightarrow$ $f(\underline{x}) \geq f(\bar{\underline{x}}) + \nabla^T f(\bar{\underline{x}})(\underline{x} - \bar{\underline{x}}) \quad \forall \underline{x}, \bar{\underline{x}} \in C$

- $f: C \to \mathbb{R}$, $f \in C^2(C)$, $C \subseteq \mathbb{R}^n$ open, non-empty convex $\Longrightarrow$ f convex $\Longleftrightarrow$ $\nabla^2 f(\underline{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)$ is p.s.d. $\forall \underline{x} \in C$

- Subgradient

  $C \subseteq \mathbb{R}^n$ convex, f convex. $\underline{\gamma} \in \mathbb{R}^n$ is a subgradient of f at $\bar{\underline{x}} \in C$ if:

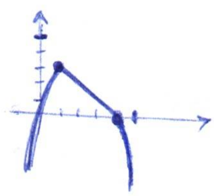
  $$f(\underline{x}) \geq f(\bar{\underline{x}}) + \underline{\gamma}^T (\underline{x} - \bar{\underline{x}}) \qquad \forall \underline{x} \in C$$

  We denote with $\partial f(\underline{x})$ the subdifferential, i.e. the set of all the subgradients of f at $\underline{x}$.

  Remark: $\underline{x}^*$ global minimum of a convex function $f: C \to \mathbb{R} \Longleftrightarrow \underline{0} \in \partial f(\underline{x}^*)$

Ex. subgradient:

$$f(x) = \begin{cases} 4 - x & 1 \leq x \leq 4 \\ 4 - (x-2)^2 & \text{otherwise} \end{cases}$$



There are two points for which f is not differentiable: $x = 1, \quad x = 4$

$\Downarrow$

$$\gamma = \begin{cases} -1 & 1 \leq x \leq 4 \\ -2(x-2) & \text{otherwise} \end{cases}$$

$x = 1: \quad \gamma = \begin{cases} -1 \\ 2 \end{cases} \Longrightarrow \partial f(1) = [-1, 2]$

$x = 4: \quad \gamma = \begin{cases} -1 \\ -4 \end{cases} \Longrightarrow \partial f(4) = [-4, -1]$